Author's Copy Educational Purposes Only

HEALTH ECONOMICS, VOL. 5: 129–140 (1996)

ECONOMICS OF PRIMARY CARE

SELECTION BIAS IN GP FUNDHOLDING

DARRIN L. BAINES AND DAVID K. WHYNES
Department of Economics, University of Nottingham, U.K.

SUMMARY

This paper uses a logistic regression model based on 1993 data for general practices in a single Family Health Services Authority (Lincolnshire) to analyse the differences in characteristics between existing fundholding (up to and including wave three) and non-fundholding practices. A high degree of classification accuracy is obtained. Fundholders are revealed to be more likely than non-fundholders to meet a number of the various quality criteria laid down by central government following the 1990 National Health Services Act, for example, with respect to prescribing cost control, minor surgery and cervical screening uptake.
The model is employed to forecast the fourth wave of fundholding and poor predictions suggest the existence of a structural break in the characteristics of fundholders between those in the first three waves and those of wave four. The evidence presented also supports the existence of selection bias in the first three waves of fundholding, although further logistic regression analysis reveals a form of such bias in the fourth wave also.

KEY WORDS—general practice; fundholding; logistic regression; NHS reforms; prescribing costs

INTRODUCTION

Fundholding by general practitioners (GPs) emerged in the UK as a result of the 1990 National Health Services Act. This Act was, in effect, a reforming strategy intended 'to raise the performance of all hospitals and GP practices to that of the best'[1] and it entailed an unprecedented increase in the economic independence of agencies within UK public sector health care. Hospitals became entitled to apply for trust status, which conferred upon them a high degree of managerial and financial autonomy. For their part, GPs who were successful in their application for fundholding status received a budget or fund, from which a range of medical services—diagnostic, inpatient and outpatient—could be purchased at the GPs' discretion. Applications for fundholding status proceeded in annual 'waves' from 1991 and recruitment to the scheme increased dramatically during the first few years. Between 1991 and 1994, the number of funds in England rose from 294 to 1,673. By the latter year, nearly one quarter of all English practices were members of the fundholding scheme, whose total budget accounted for around 9 per cent of total National Health Service (NHS) resources.[2]

Surveys conducted during the earliest waves of fundholding have suggested that applicants were anticipating a wide range of benefits from their new status, including improved management and computer services, a more accessible service from local hospitals and freedom from administrative restrictions.[3,4] Theoretically, there existed efficiency gains which fundholders ought to have been able to reap, for example, increased responsiveness on the part of providers.[5] However, the extent to which these potential gains have been realised, and at what cost, remains a moot point, owing to the limited extent of research in the field. Much of that which *does* exist is anecdotal rather than scientific. The question is particularly significant, of course, in view of the current policy initiative towards 'total purchasing',

Address for correspondence: D. K. Whynes, Department of Economics, University of Nottingham, Nottingham NG7 2RD, U.K. Tel: 0115 951 5463; Fax: 0115 951 4159.

CCC 1057–9230/96/020129–12
© 1996 by John Wiley & Sons, Ltd.

Received 1 May 1995
Accepted 24 January 1996

whereby the GPs in a locality purchase *all* hospital and community health services for their patients.[6]

EVALUATING FUNDHOLDING

The success or otherwise of the fundholding experiment has proved difficult to evaluate thus far, for four main reasons. First, the 1990 NHS Act was implemented in the total absence of any formally-established structures of monitoring and evaluation. Moreover, research into the progress of the reform process conducted by researchers outside government received little official support. Second, the Act itself was accompanied by other substantial changes in public health care provision, which themselves impacted upon NHS development. Examples include the 1990 revisions to the GP contract, changes in the management of community care, the 'Patient's Charter' initiative and a sharp increase in the volume of real resources allocated to the NHS.[7] It is accordingly conceptually difficult to assign an observed phenomenon to any one of these possible causes. Thus, although studies of first-wave GP fundholders have suggested they appear to have been more successful than non-fundholders in controlling their prescribing budgets and were proving more adept than health authorities in contract negotiation with care providers,[8-10] other researchers have questioned whether those practice-based innovations observed are necessarily a consequence of fundholding.[11]

Third, it was noted above that the overt intention of the 1990 NHS Act, and thus one of the goals of fundholding, was to raise the performances of all GP practices to those of the best. However, the definition of best, or, for that matter, merely good, practice in primary care has never been formally agreed. This comes as no surprise; there are probably at least as many definitions of good practice as there are GPs. Evaluation in the absence of an agreed objective function is clearly unworkable. This having been said, it *is* possible to employ the various pronouncements and policy initiatives of central government to make reasonable inferences about how the management of the NHS is interpreting quality in general practice. Following the 1990 Act, the Family Health Service Authorities (FHSAs), which were responsible for the management of general practice, were provided with a clear development agenda. Within this agenda, the clinical issues to be addressed included controlling prescribing costs, developing health promotion, implementing new target payments schemes and establishing arrangements for minor surgery. Non-medical staff establishments within practices were to be strengthened. List sizes below 2,500 per GP at the FHSA level became an important performance indicator, on the grounds that GPs with longer lists appeared less successful in meeting many of the desired practice objectives.[12-14]

The final problem faced in evaluating fundholding involves the possible existence of selection bias. By way of illustration, consider the case of hospital trusts. In common with fundholding, trust status is optional and those hospitals who applied initially were clearly self-selecting. Suppose that those who actively sought the new status in the first wave were distinguished by the possession of some common characteristic which pre-disposed them towards application. Evidence indeed suggests this to have occurred, the common characteristic being relatively high levels of efficiency.[15] However, there is no guarantee that subsequent waves of trusts would necessarily share the characteristics of those in the earlier waves, with the result that the aggregate performance of trusts might begin to change. Thus, our assessment of the virtues or otherwise of trust status might require revision with continued recruitment.

In this paper, we explore the possibilities of a form of selection bias amongst fundholding general practices within a single FHSA. First, we examine a set of data to identify possible differences in characteristics, widely defined, between existing fundholders (FHs) and non-fundholders (NFHs) at a particular point in time. We then develop a statistical model to classify the FHs and NFHs according to criteria derived from these characteristic differences, and establish whether these are characteristics associated with practice quality as officially defined. Finally, we employ this model to predict which of the existing NFHs are likely to become FHs in future waves and compare our predictions with known events. This permits us to determine whether subsequent waves of FHs (i) are simply those NFHs possessing the same characteristics as existing FHs, only to a lesser degree, or (ii) are qualitatively different, implying that a structural break in recruitment has occurred. If the answer is case (i), we have evidence of selection bias as defined earlier. The behaviour paths of FHs and NFHs will continue to

diverge as the relevant characteristics are selected out of the NFH group. If the answer is (ii), the characteristic differences between FHs and NFHs will begin to break down and the perceived outcomes of the NHS reforms with respect to general practice as a whole will start to change.

EXPLORING DIFFERENCES IN CHARACTERISTICS

The data used in this investigation pertain to practices in the Lincolnshire FHSA for calendar year 1993. At this time, there existed 108 general practices in Lincolnshire, of which 19 had attained fundholding status in one of the first three waves. Eight NFHs and one FH had to be excluded from the analysis owing to incomplete data, giving a total sample of 99 practices, comprising 81 NFHs and 18 FHs. Ideally, it would have been desirable to analyse change on a wave-by-wave basis, although this was precluded by the small numbers of fundholding practices in the individual waves (2, 6 and 10 in waves 1 to 3, respectively).

Table 1 presents means and standard deviations for a wide range of practice characteristics, on the basis of fundholding status (NFHs vs. FHs). These characteristics pertain both to the practices themselves and to the lists of patients which they serve. Also included in the Table are significance levels (P-values) for differences in the means for each characteristic. For the continuous variables, these were derived from the bivariate t-test, which is known to be robust against departures from normality. For the binary variables, the Mann-Whitney test was employed.[16] These data should be viewed primarily as exploratory statistics, to identify relevant variables for inclusion in the classification model. Under the Bonferroni criterion governing the presentation of multiple significance test results, we should require a superior level of significance to the normal five per cent level in order to accept a difference in means for each individual comparison.[17] In our case, the precise level of significance is difficult to estimate with precision in view of high degrees of multicollinearity within sub-sets of the data.

To facilitate exposition, the variables have been grouped loosely under six headings. The first class includes characteristics of the membership of practice lists. Socio-demographic factors, such as age and sex composition, are represented, as are proxies for practice morbidity and social deprivation, for example, rates of permanent sickness and unemployment. Second, data are presented for geographical or locational variables, namely, the degree of proximity of other practices and hospitals. In addition, the average rural practice payment, which is calculated on the basis of patient-surgery distances,[18] represents the spatial dispersion of the practice list. The variables in this second class are essentially measures of urbanisation, or the lack of it. Lincolnshire, it should be borne in mind, is a predominantly rural county with a limited number of urban centres. The P-values in excess of 0.05 for mean differences for all the characteristics in these two classes suggest that hypotheses to the effect that the FHs in the first three waves were drawn from practices with particular geographical or patient list characteristics are unlikely to be sustainable, especially in view of the Bonferroni criterion.

The third class of variables relate to practice organisation, including referrals to hospitals. The dispensing practice variable was included because, owing to the rural nature of the county, Lincolnshire contains a high proportion of such practices (in excess of 50 per cent). Dispensing practices are permitted to dispense their own prescriptions and have been shown to generate higher prescribing costs.[19] However, our data provide no evidence that the representation of dispensing practices is higher or lower amongst FHs. The proposition that FHs are more likely than NFHs to be training practices, however, receives some support. Mean list-size is evidently vastly different between FHs and NFHs, although this is to be expected in view of the list-size requirements which were originally imposed upon applicants to fundholding. The eligibility requirements for waves 1 through 3 were 11,000, 9,000 and 7,000, respectively.[10] Beyond this difference, there is reason to suspect that FHs recorded lower hospital referral rates and operated fewer branch surgeries. Whilst some NFHs had an average list per GP in excess of 2,500 patients, none of the FHs fell into this category. This particular average list-size figure is significant as it has emerged as a general practice performance indicator.[20]

The fourth class of variables pertain to prescribing and here the case for differences between the means appears strong. FHs appear to be lower-cost prescribers, both per member of the list and per ASTRO-PU, the latter being a formula which translates patients into 'prescribing units' by incorporating weightings based on age, sex and

Table 1. Exploratory statistics of 46 characteristics, 81 NFHs vs. 18 FHs, 1993

	NFH		FH		
Characteristic	Mean	SD	Mean	SD	P=
List					
List inflation owing to temporary residents (%)	6.44	12.91	9.13	18.89	0.561
Night visits per 1000 patients per quarter	33.68	10.07	34.85	8.39	0.649
Proportion of list aged ⩾ 65 years (%)	19.48	5.00	18.08	2.25	0.074
Proportion of list bearing pre-payment certificates (%)	3.69	0.92	3.57	0.45	0.606
Proportion of list defined as permanently sick (%)	3.80	1.14	3.42	1.04	0.196
Proportion of list resident in nursing homes (%)	1.35	1.43	1.00	0.72	0.139
Proportion of list which is female (%)	51.15	1.89	51.26	0.73	0.681
Unemployment rate (%)	8.29	3.41	7.61	2.95	0.435
UPA8 (Jarman score)	−2.00	11.49	−2.48	7.99	0.834
Location					
Miles from practice to nearest hospital	6.98	5.04	7.24	6.54	0.853
Number of other practices within a one mile radius	1.46	2.02	1.56	1.50	0.845
Rural practice payment per patient	0.26	0.22	0.23	0.16	0.415
Proportion of list living >3 miles from practice (%)	27.86	23.68	22.86	15.54	0.275
Organisation					
Dispensing practice? (1 = Yes, 0 = No)	0.60	0.49	0.44	0.51	0.215
Training practice? (1 = Yes, 0 = No)	0.21	0.41	0.44	0.51	0.039
Less than 2,500 patients per GP? (1 = Yes, 0 = No)	0.83	0.38	1.00	0.00	0.058
List (patients) per GP	1913.10	574.08	1961.26	217.49	0.558
List-size growth during year (%)	0.44	8.26	1.04	1.46	0.760
List-size/1000	4.63	2.80	10.63	3.16	0.000
Number of branch surgeries operated	0.89	1.20	0.22	0.43	0.000
Hospital referrals per 1000 patients on list	172.19	58.08	150.91	31.40	0.035
Ratio of emergency to elective hospital admissions (%)	19.69	4.79	19.24	4.29	0.716
Home visits per patient per year	0.41	0.25	0.52	0.26	0.108
Ratio of day case to ordinary admissions (%)	19.69	4.79	19.24	4.29	0.716
Number of hospitals to which patients referred	8.52	6.55	14.83	7.45	0.000
Prescribing					
Prescribing cost per ASTRO–PU (£)	19.33	3.24	16.96	2.13	0.004
Prescribing cost per patient (£)	81.63	16.53	68.71	11.91	0.007
Prescribing cost per patient in 1990 (£)	56.61	12.98	48.66	9.16	0.016
Prescribing budget overspend (% of budget)	7.02	10.72	1.48	4.63	0.035
Prescribing budget per ASTRO–PU (£)	18.82	3.35	15.75	1.99	0.000
Proportion of medicines prescribed generically (%)	34.82	12.61	43.18	14.49	0.015
Practice receives PACT level 3 information? (1 = Yes, 0 = No)	0.35	0.48	0.61	0.50	0.038
Structure					
Age of senior partner (years)	49.94	8.42	54.61	5.12	0.004
Average age of partners (years)	44.38	6.65	42.94	2.46	0.130
Number of partners	2.51	1.36	5.44	1.58	0.000
Nurse hours per GP	15.97	10.39	14.38	5.63	0.531
Practice/fund manager hours per GP	14.75	11.19	12.60	4.59	0.196
Proportion of partners which are female (%)	21.33	26.45	14.58	11.60	0.098
Targets and income					
Cervical cytology uptake (% of target)	82.49	8.38	86.02	3.00	0.003
Child health surveillance (% of target)	83.02	30.14	91.20	9.28	0.044
Health promotion clinics per 1000 patients	16.24	15.44	19.09	8.37	0.953
Episodes of paid minor surgery per GP	9.07	4.58	11.43	1.06	0.000
Pre-school surveillance uptake (% of target)	88.88	15.73	95.44	4.34	0.002
Under-2 child immunisation rate (% of target)	96.60	3.54	96.70	2.71	0.911
Maternity claims per 1000 patients	3.43	1.68	3.01	0.75	0.300
Vaccinations per 1000 patients	128.11	67.41	134.88	49.14	0.688

temporary resident status.[21] Because the prescribing budget was generally set, in part, on the basis of historic cost figures, FHs' 1993 prescribing budgets appear to be lower than those of NFHs, yet their mean overspend was also lower. A higher proportion of FH's prescribed medication was generic and a higher proportion of FHs subscribed to the most detailed 'Prescribing Analysis and CosT' information (PACT Level 3) which has been available to all GPs since the late-1980s.[22] Both of these factors presumably contribute to FH cost control although it should be noted that such practices exhibited lower prescribing costs in 1990, prior to the introduction of fundholding.

The next class of variables relates to practice structure. The mean size of fundholding partnership is evidently much larger than that of NFHs; again, this is to be expected in view of the list-size requirement for eligibility. In addition, it emerges that the senior partner of FHs was older than that of NFHs. There is no evidence that FHs employed a higher or lower level of management input, nor that the decision to fundhold was influenced by the gender of partners. The final class of variables relates to practice activities involving the achievement of targets and income generation. FHs evidently possess the better record with respect to cervical cytology uptake, child health and preschool surveillance, although no differences in performance can be detected in other areas. FHs appear to undertake a higher amount of paid minor surgery.

PREDICTING FUNDHOLDERS

In developing our classification model, we utilised the model building strategy advocated by Hosmer and Lemeshow.[23] This entails the pre-selection of potential explanatory variables on the basis of significance testing, although using a more liberal criterion than the 5 per cent significance level. We considered that variables in Table 1 with a significance of difference in means at $P = 0.2$ or better were candidates for inclusion in the classification analysis. Fisher's discriminant analysis is often used in classification exercises such as this but requires that the variables in each group are from multivariate normal distributions and that the covariance matrices for all groups are equal.

Investigation of our data revealed that these assumptions did not always hold. As 'it is difficult to say how much flexibility can be granted before the method becomes unreliable',[24] we opted for a classification method less vulnerable to departures from the standard assumptions, namely, logistic regression. Logistic regression differs from linear regression in that the dependent variable is dichotomous (in our case, 1 if FH and 0 if NFH). The regression result produces, in effect, an estimate of the probability that a practice with given values of characteristics is fundholding or otherwise. If the estimated coefficient, β, is positive, then a practice is more likely to be fundholding if higher levels of the characteristic associated with that coefficient are present. Where β is negative, more of that characteristic diminishes the probability of being a fundholder. A useful property of the estimation is that e^{β} represents the change in the odds ratio for a marginal change in the independent variable.

One further practical problem remains with respect to modelling from our pre-selected data set. A minimum list-size, it will be recalled, was a requirement for attaining fundholding status in each of the first three waves. Accordingly, a characteristic common to all our 18 FHs is a list-size above 7,000. Of the 81 NFHs, only 10 NFHs possessed list sizes above the statutory minimum and were therefore, in principle, eligible to join the scheme. A classification based simply upon list-size eligibility will correctly classify 100 per cent of FHs and 87.7 per cent of NFHs, with an overall correct classification of 89.9 per cent. This represents a high degree of classification accuracy but imparts, of course, no new information. There are, in principle, three ways of accommodating the bias resulting from the list requirement. First, we might attempt to classify a sub-set of cases, actual FHs and those NFHs eligible to apply for fundholding. However, this results in an unsatisfactorily small sample of 28 cases. Second, one might want to include eligibility in the classification model as a dummy variable. However, as is evident from the results above, the effect of eligibility is so strong that all other discriminating factors are swamped. Again, no new information is imparted. The third possibility, and the one eventually adopted, was to exclude list-size from our list of pre-selected variables. In addition, we excluded the number of partners, which was highly correlated with it ($r = 0.94$). Our classification is thus intended to identify discriminating variables, other than those imposed externally.

The logistic regression model was estimated using the backward stepwise procedure. The

criteria for overall goodness-of-fit in logistic regression are less straightforward, and possibly more subjective, than for linear regression. Whilst the statistical significance of the estimated coefficients can be estimated in precisely the same manner as for linear regression, namely, T-ratios >1.96 for significance at the 5 per cent level, there is no logistic equivalent to the coefficient of determination. Instead, it is conventional to report 'model improvement', the difference between the log-likelihood function of a model including only the constant term and that of the model with both the constant and relevant variables included. The model improvement chi-squared tests the null hypothesis that the coefficient of an excluded variable in the final model is zero, and is analogous to the F-change test in linear regression. The degree of classification accuracy naturally provides supporting information regarding a model's goodness-of-fit.[23] It should be noted that collinearity was a particular problem in this data set, especially amongst the prescribing variables. Accordingly, a permutated series of models was tested, in which each of the collinear variables appeared independently of its correlates.

Table 2 presents the logistic model finally chosen. As may be inferred, it correctly classifies all but 3 FHs and 3 NFHs. All coefficients achieve statistical significance at the 5 per cent level and the significances are robust against the inclusion of any additional variables from the pre-selected list. Pearson correlation coefficients between each of the final model's variables and list size never exceed 0.3, providing reason to believe that the model is reasonably uncontaminated by the statutory size condition. The three NFHs predicted as FHs possessed list sizes of around 5,000, and were thus ineligible for fundholding during the first three waves. The three FHs predicted as NFHs all possessed list-sizes in excess of 10,000.

The odds ratios presented in Table 2 indicate the chances of being classified as a FH, for marginal increases in the independent variables. Odds ratios above 1 clearly enhance the chances of a practice being classified as a FH, whilst those less than 1 diminish the chances. Thus fundholding status is *more probable* when the age of the senior partners, target fulfilment for cervical cytology and the rate of paid minor surgery are high. On the other hand, fundholding status is *less likely* where the number of branch surgeries operated, the mean ages of the partners, overspend on the prescribing budget and the ratio of emergency to elective admissions are high.

Classification based on statistical association is important in demonstrating where characteristic differences lie but requires an appeal to intuition or theory before the results become acceptable as a realistic model. To what extent do the signs on these coefficients accord with such intuition? To begin with, the finding that the senior partner of a fundholding practice is typically older than his/her

Table 2. Logistic regression, 18 FHs vs. 81 NFHs

Variable	β	SE	P=	Odds ratio
Age of senior partners (years)	0.57	0.18	0.00	1.77
Average age of partners (years)	−0.79	0.30	0.01	0.45
Cervical cytology uptake (% of target)	0.31	0.13	0.02	1.37
Number of branch surgeries	−2.30	1.01	0.02	0.10
Prescribing budget overspend (% of budget)	−0.29	0.10	0.00	0.75
Episodes of paid minor surgery per GP	0.80	0.38	0.04	2.23
Ratio of emergency to elective admissions (%)	−0.09	0.04	0.02	0.92
Constant	−24.42	12.77	0.06	
Goodness-of-fit				
Initial log-likelihood function	93.88			
Final log-likelihood function	33.96			
Model improvement (Chi-squared)	59.92	(P = 0.00)		
Correct classification (%)				
Fundholders	83.33			
Non-fundholders	96.30			
Overall	93.94			

NFH counterpart is *not* unexpected as this has emerged elsewhere in ongoing research. Survey evidence suggests that some senior GPs have seen fundholding essentially as a means of job enrichment, in that it adds a major new dimension to a pattern of work which is otherwise unchanging over time.[25] In addition, the older senior partners are, of course, the closest to retirement. As fundholding permits the re-investment of budget surpluses in practice improvement, and thereby increases the capital (hence resale) value of the practice premises, we should anticipate that older senior principals would be more conscious than their younger counterparts of using fundholding status as a means of leaving the service on the most favourable financial terms possible.

Were the signs on the coefficients for cytology uptake, paid minor surgery and prescribing budget overspend to be the reverse of those identified, we should be obliged to conclude that FHs were perverse in their responses to the institutional structures created by the fundholding scheme and by the 1990 GP contract. There would appear to be no grounds for accepting this conclusion *a priori*. In contrast, however, the finding that FHs are characterised by a lower ratio of emergency to elective admissions *does*, at first sight, appear perverse. Given that emergency admissions are extra-budget, we might anticipate that FHs would attempt to move a higher proportion of their patients to the emergency category. This particular incentive argument is naive, however. In the first place, such a strategy would only represent a short-run optimum to the FH as he/she ought to anticipate the FHSA revising his/her budget downwards in the following year, based on this year's elective contracts. Second, the FH is responsible for his/her own contract placements and, as is evident from Table 1, regularly uses a far higher number of providers than does the NFH working through the FHSA. The sign on the coefficient is thus consistent with the evidence cited earlier, to the effect that FHs generally have been more successful than NFHs in placing elective admission contracts. The latter, presented with more restricted choice of provider, face the reverse incentive, namely, to classify more patients as emergencies in order to gain their admission.[26]

The explanation of the importance of the number of branch surgeries is more problematic, as it remains unclear what motivates a practice to run such surgeries. Within our data set, bivariate correlation between this variable and all others reveals that locational effects offer the best explanation: correlation coefficients between the number of branch surgeries and rural practice payment or proportion of list living more than three miles distant are 0.37 and 0.44, respectively. No other comparison produces a coefficient in excess of 0.25. This having been said, it is clearly evident that the factors encompassed by our data set fail to capture a substantial proportion of the explanation, and the decision to run branch surgeries might therefore depend more on the professional attitudes of the members of individual practices, in particular, attitudes towards 'proper' service to patients. In the past, such an explanation has been offered to account for otherwise unpredictable variations in the rate of GP night visiting.[27,28]

PREDICTING THE FOURTH WAVE

Our analysis was conducted on data for 1993. At this time, 21 of the NFH practices in our sample were in their preparatory year, prior to the achievement of full independent status as fourth-wave fundholders in 1994. All were represented in our data set. Although data for a single year are not completely robust in describing a future state of the world following a change of status, it is pertinent to ask—how well does the model derived above predict the fundholders of the next wave? There are two ways of answering this question.

First, we can simply note that none of the fourth wave potential FHs were predicted as actual FHs by our model of Table 2. Elaborating slightly, we can utilise the probability scores of the 81 NFHs estimated from the logistic regression to compare the average probability of fundholding for the 21 fourth-wave FHs and the remaining 60 NFHs. For the former, the mean probability of being a fundholder was 0.053 and, for the latter, 0.070. These probability distributions are highly skewed (skew coefficients = 1.636 and 3.218, respectively) with the result that the Mann-Whitney test is appropriate, in order to determine whether the differences in the averages are significant. This test yields the result $Z = -1.013$, $P = 0.311$. Accordingly, fourth-wave fundholders appear to have no statistically significant higher chance of being classified as FHs by the original logistic regression model than do NFHs.

Second, it is possible to re-run the logistic regression, starting with the variables as specified

in Table 2, with only the 81 NFH practices of 1993, this time with a view of discriminating between the 21 fourth-wave FHs and the 60 NFHs. After stepwise elimination, significant coefficients remain only for average age of partners, episodes of paid minor surgery and number of branch surgeries operated (but with the coefficient sign reversed from Table 2). This formulation yields a model improvement of 14.976 on an initial log-likelihood function of 92.709. It correctly predicts 5 future FHs (23.8 per cent), whilst mis-classifying 4 NFHs as FHs. It should be noted, in addition, that none of the three NFHs in the first three waves predicted as FHs by the Table 2 model actually became fundholding in the fourth wave. These results suggest that the model developed to discriminate between FHs and NFHs in the first three waves performs poorly with respect to predicting the fourth wave. This leads to the conclusion that the distinguishing characteristics of fourth-wave FHs prior to fundholding must, at least in some degree, be different from those of existing FHs. In order to discover what these new characteristics are, it is necessary to repeat the model-building procedure described above, specifically for the 1993 NFH sub-sample of 21 fourth-wave potential FHs and the remaining 60 NFHs. However, one important difference in our data requirements now emerges. Practices electing to fundhold in wave four and beyond were no longer governed by a strict list-size requirement. Instead, they were permitted to establish multi-funds, groups of practice with a combined list size in excess of 7,000. Accordingly, the list-size and partnership size variables need no longer be excluded from the pre-selected data set. In fact, the list-sizes of fourth-wave FHs were normally distributed about a mean of 5,182 (SD = 2,453), with range extremes at 1,100 and 10,800.

Table 3 presents the final logit model chosen. Examining the odds ratios, it appears that fourth-wave fundholding status is *more probable* where list-sizes are larger and practice manager hours per GP are higher. Fundholding status is *less likely* where the preponderance of elderly patients in the list, the average age of the partnership and the rate of hospital referrals are higher. It should be noted that none of the prescribing variables listed in Table 1 appeared in the logistic model for the fourth wave; indeed no significant differences in means between fourth-wave FHs and the remaining NFHs (at the 5 per cent level, even without the Bonferroni correction) were observed for *any* of the prescribing variables.

Our results suggest that forms of selection bias were in operation at the fourth wave of fundholding. In spite of the fact that the formation of practice consortia to reach the statutory list-size was permitted,[29] individual practice size was an important characteristic of the fourth-wave fundholding recruits. This would seem to imply that economies of scale in fundholding were perceived to exist (possibly fundholding was considered too difficult to manage in small practices). Alternatively, it has been demonstrated that random fluctuations in practice morbidity are more severe

Table 3. Logistic regression, fourth-wave FHs vs. NFHs

Variable	β	T-ratio	Odds ratio
Proportion of list aged ≥ 65 years (%)	−0.457	−2.909	0.633
Hospital referrals per 1000 list	−0.029	−2.344	0.971
Practice manager hours per GP	0.241	3.410	1.273
Average age of partners (years)	−0.361	−2.557	0.697
List-size/1000	0.402	2.608	1.496
Constant	21.138	2.761	
Goodness-of-fit			
Initial log-likelihood function	92.709		
Final log-likelihood function	42.906		
Model improvement (Chi-squared)	49.803	(P = 0.00)	
Correct classification (%)			
Fundholders	76.19		
Non-fundholders	95.00		
Overall	90.12		

with smaller list sizes.[30] The concomitant financial uncertainty might therefore have contributed to the decision of smaller practices not to apply for fundholding status.

The high level of management input characterising the newer FHs might be interpreted as a response to a desire to improve efficiency to the level of existing FHs. Because the fourth-wavers in their preparatory year were able to claim 50 per cent of the full fundholder's management allowance, the cost of their disproportionate management input was underwritten in some degree. The finding that fourth-wavers had fewer elderly patients on their list and lower rates of hospital referral is consistent with the view that the morbidity characteristics of their patients was more favourable than that of the remaining NFHs.

Only the average age of the partnership appears common to both classification models and, in both cases, the sign of the coefficient is the same. The models suggest that younger partnerships are more likely to be fundholders, and this in spite of the fact that existing FHs have older senior partners. Of all the variables included, it is average age which is most difficult to explain on a theoretical or intuitive basis. Whilst one might retreat into folklore and suggest that 'younger practices' are less likely to be conservative in their thinking, we feel that a full explanation for the significance of this variable lies well beyond the possibilities of our data set. A comprehensive answer would require, *inter alia*, a detailed audit of the patterns of partner recruitment and of the nature of decision making within the individual practices.

HAS SELECTION BIAS OCCURRED?

The results of the two logistic analyses appear to support the existence of a substantive structural break in the process of recruitment to fundholding in Lincolnshire between the third and fourth waves. Over and above their fulfilment of the list size criterion, the distinctive characteristics of Lincolnshire FHs up to and including wave three were a higher rather than a lower paid minor surgery caseload, higher rather than lower rates of uptake for cervical screening, older senior partners and a lower average age for the partnership, financial discipline with respect to the prescribing budget, lower ratios of emergency to elective hospital admission and fewer branch surgeries. This set of characteristics is at variance with that which describes the fourth-wave of FHs. Whilst list size is important for these FHs, this characteristics set comprises a higher proportion of younger patients, a lower rather than higher rate of hospital referral and a higher rather than lower utilisation of practice management resources.

Our analysis of FHs in the first three waves (Tables 1 and 2) supports the view that these were practices who already possessed the capacity, or possessed the ability to develop the capacity in a very short space of time, to best meet most of the official quality criteria laid down. This is especially evident with respect to prescribing cost control, screening and surveillance targets, minor surgery and list-size per GP. It is vital to appreciate that none of these characteristics which discriminate FHs from NFHs are necessary consequences of fundholding status. Our analysis therefore strongly supports the existence of selection bias in the earlier waves of fundholding, namely, that those practices obtaining fundholding status in the early waves were those most capable of achieving success as officially assessed.

Does the existence of selection bias make a difference to the quality of the services provided by the practices? Table 4 provides a comparison between the existing (waves one through three) FHs, fourth-wave FHs and remaining NFHs with respect to a number of the success indicators included in the FHSA agenda. Significance levels relate to the Kruskal-Wallis one way analysis of variance test, which was preferred to its parametric equivalent on the grounds that median values are likely to be of more interest than means in distributions exhibiting some degrees of skew.[31] In addition, it should be noted that all of the existing FHs possessed list-sizes per GP under 2,500, whereas the proportions for fourth-wave FHs and NFHs were 81 and 83 per cent, respectively. With respect to target payments and health promotion, differences between the three groups appear statistically insignificant (although the health promotion activities of fourth-wave FHs appears noticeably lower than for the other groups). In that these are income-generating services, a lack of difference is possibly unsurprising, although it should be borne in mind that the cost effectiveness of most of these services (and thus the consequences for patients' health) is unproven.[32]

From Table 4, the statistically-strongest relationships emerge with respect to prescribing costs. Here there are good grounds for accepting that existing FHs are the lowest-cost prescribers and

Table 4. Differences in medians for key performance characteristics, 1993

Characteristic	WAVES 1–3	WAVE 4	NFH	P=
Prescribing cost per ASTRO–PU (£)	17.01	19.32	19.18	0.01
Prescribing budget overspend (% of budget)	2.42	5.82	7.30	0.01
Proportion of medicines prescribed generically (%)	41.23	33.90	35.25	0.07
Health promotion clinics per GP	17.29	7.98	14.36	0.33
Cervical cytology uptake (% of target)	86.60	84.92	83.82	0.23
Child health surveillance uptake (% of target)	94.34	96.21	96.34	0.68
Pre-school surveillance uptake (% of target)	94.42	89.91	91.75	0.09
Under-2 child immunisation rate (% of target)	97.34	96.77	97.38	0.77
Episodes of paid minor surgery per GP	11.90	11.80	10.83	0.09

are more successful at remaining within budgetary limits. In that generic medicines are presumed to be cheaper than branded equivalents,[33] it is not unreasonable to ascribe a portion of Lincolnshire FHs' economies to their higher use of generics. The full-sample correlations between the use of generics, budget overspends and prescribing costs per patient, all of the order of −0.35, supports this belief, as do findings reported in other studies.[34] It is debatable whether cheap prescribing is also good prescribing[35] and our data set provides few further clues in this respect. It should be noted that our findings at this stage lend no support to the hypothesis that prospective fundholders will increase their current-period prescribing costs in order to inflate the value of their eventual prescribing budget.[36] In fact, they mirror the findings of a recent Scottish study.[37] Although an economic incentive for such an increase might appear to exist, countervailing factors will be present. For example, high-cost prescribers would not present themselves to the health authority as particularly suitable candidates for fundholding and any unusual 'cost-hikes' in the preparatory year would be clearly visible. As Table 4 indicates, the fourth-wave FHs did not typically possess above-average prescribing costs pre-fundholding although they *did* possess considerable potential for cost reductions post-fundholding, in view of their low current rate of generic prescribing.

Finally, we must consider the extent to which the results of our analysis are generalizable beyond Lincolnshire. If only by virtue of the facts that the Lincolnshire FHSA records the highest prescribing cost per patient in England[38] and that it is one of the most rural of English counties, the area can hardly be considered typical. This having been said, it is not obvious that the characteristics identified in our classification models, with the possible exception of branch surgeries, are intuitively irrelevant as potential explanatory factors in other regions. Obviously, a definitive answer to the generalizability question will only emerge when more locality studies are published.

According to the rhetoric of the 1990 NHS changes, fundholding was intended to facilitate efficiency gains. Whether or not such efficiency gains will be realised remains an open question and one which cannot be answered with any degree of certainty until several years hence. Whilst it would not seem implausible to suppose that increased management input of the newer FHs could contribute to superior performance, the issue of prescribing remains more problematic. Both the data in Table 1 and our earlier study of prescribing behaviour[39] strongly suggest that early-wave FHs' control over prescribing costs was an inherited characteristic, rather than one acquired as a result of fundholding. If prescribing costs indeed prove difficult to control in the later waves of FHs, then it follows that the comparative past success of the fundholding scheme with respect to prescribing, as noted in the earlier studies, is indeed an artefact of selection bias. Fundholders' prescribing performance as a whole will, in the future, regress towards that on the non-fundholders.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the illuminating comments of the anonymous referees, the statistical guidance of Professor Paul Newbold, and the financial and research support of Lincolnshire Health. The opinions expressed in this paper are those of the authors alone and do not necessarily represent those of Lincolnshire Health.

REFERENCES

1. CM 555 *Working for patients*. London: HMSO, 1989.
2. Comptroller and Auditor General. *General practitioner fundholding in England*. National Audit Office, 1994; HC 51 Session 1994–95.
3. Bain, J. Budget holding: here to stay? *British Medical Journal* 1993; **306**: 1185–1188.
4. Newton, J., Fraser, M., Robinson, J. and Wainwright, D. Fundholding in Northern region: the first year. *British Medical Journal* 1993; **306**: 375–378.
5. Glennerster, H. The future of fundholding. In: Harrison, A. (ed) *Health care UK, 1993/94*. London: King's Fund Institute, 1994: 45–49.
6. NHS Management Executive. *Developing NHS purchasing and GP fundholding*. Department of Health, Leeds, 1994; EL(94)79.
7. Le Grand, J. Evaluating the NHS reforms. In: Robinson, R. and Le Grand, J, (eds) *Evaluating the NHS reforms*. London: King's Fund Institute, 1994: 243–260.
8. Bradlow, J. and Coulter, A. Effect of fundholding and indicative prescribing schemes on general practitioners' prescribing costs.*British Medical Journal* 1993; **307**: 1186–1189.
9. Glennerster, H., Matsaganis, M., Owens, P. and Hancock, S. *Implementing GP fundholding*. Buckingham: Open University Press, 1994.
10. Glennerster, H., Matsaganis, M., Owens, P. and Hancock, S. GP fundholding: wild card or winning hand? In: Robinson, R. and Le Grand, J. (eds) *Evaluating the NHS reforms*. London: King's Fund Institute, 1994: 74–107.
11. Corney, R. Experiences of first wave general practice fundholders in South East Thames Regional Health Authority. *British Journal of General Practice* 1994; **44**: 34–37.
12. Audit Commission. *Practices make perfect: the role of the Family Health Services Authority*. London: HMSO, 1993.
13. Huntington, J. From FPC to FHSA to ... health commission? *British Medical Journal* 1993; **306**: 33–36.
14. NHS Management Executive. *Health service indicators handbook*. London: Department of Health, 1993.
15. Bartlett, W. and Le Grand, J. The performance of trusts. In: Robinson, R. and Le Grand, J. (eds) *Evaluating the NHS reforms*. London: King's Fund Institute, 1994: 54–71.
16. Armitage, P. and Berry, G. *Statistical methods in medical research*. (3rd. ed.) Oxford: Blackwell Scientific Publications, 1994.
17. Bland, J. M. and Altman, D. G. Multiple significance tests: the Bonferroni method. *British Medical Journal* 1995; **310**: 170.
18. Ellis, N. and Chisholm, J. *Making sense of the Red Book*. (2nd. ed.) Oxford: Radcliffe Medical Press, 1993.
19. Morton–Jones, T. J. and Pringle, M. A. L. Prescribing costs in dispensing practices. *British Medical Journal* 1993; **306**: 1244–1246.
20. Salinas–Jimenez, J. and Smith, P. *Data envelopment analysis applied to quality in primary health care*. University of York, Centre for Health Economics, 1994; Discussion Paper 124.
21. Roberts, S. J. and Harris, C. M. Age, sex and temporary resident originated prescribing units (ASTRO–PUs): new weightings for analysing prescribing of general practices in England. *British Medical Journal* 1993; **307**: 485–488.
22. Heywood P. PACT and prescribing. *The Practitioner* 1991; **235**: 168–169.
23. Hosmer, D. W. and Lemeshow, S. *Applied logistic regression*. New York: John Wiley and Sons, 1989.
24. Altman, D. G. *Practical statistics for medical research*. London: Chapman and Hall, 1991.
25. Ennew, C. T., Feighan, T. and Whynes, D. K. Entrepreneurial behaviour in the public sector: the case of GP fundholders. Paper presented at the British Academy of Management conference: Revitalising organisations–the academic contribution. University of Sheffield, 11–13 September, 1995.
26. Court, C. Rising emergency admissions disrupt NHS. *British Medical Journal* 1994; **309**: 1322.
27. Cubitt, T. and Tobias, G. Out of hours calls in general practice: does the doctor's attitude alter patient demands? *British Medical Journal* 1983; **287**: 28–30.
28. Usherwood, T. P., Kapasi, M. A. and Barber, J. H. Wide variations in the night visiting rate. *Journal of the Royal College of General Practitioners* 1985; **35**: 395.
29. Trent Health. *GP fundholding in practice*. Sheffield: Trent Health, 1993.
30. Crump, B. J., Cubbon, J. E., Drummond, M. F., Hawkes, R. A. and Marchment, M. D. Fundholding in general practice and financial risk. *British Medical Journal* 1991; **302**: 1582–1584.
31. Siegel, S. and Castellan, N. J. *Non-parametric statistics for the behavioural sciences*. (2nd. ed.) New York: McGraw–Hill, 1988.
32. Scott, T. and Maynard, A. K. *Will the new GP contract lead to cost effective medical practice?* Centre for Health Economics, University of York, 1991; Discussion Paper 82.
33. Audit Commission. *A prescription for improvement: towards more rational prescribing in general practice*. London: HMSO, 1994.
34. Coulter, A. Evaluating general practice fundholding in the United Kingdom. *European Journal of Public Health* 1995; **5**: 233–239.
35. Griffin, J. P. *Cheap prescribing—can we afford it?* London: Association of the British Pharmaceutical Industry, 1994.
36. Glennerster, H., Matsaganis, M. and Owens, P. *A foothold for fundholding*. London: King's Fund Institute, 1992.

37. Healey, A. T. and Reid, A. P. Do prospective fundholders inflate their prescribing costs? A study of Grampian fundholding and non-fundholding practices. *Health Bulletin* 1994; **52**(4): 282–284.
38. PPA. *Prescription Pricing Authority Annual Report, 1992–93*. Newcastle upon Tyne: PPA, 1993.
39. Whynes, D. K., Baines, D. L. and Tolley, K. H. GP fundholding and the costs of prescribing. *Journal of Public Health Medicine* 1995; **17**: 323–329.